Writing an Ethical Impact Statement for ACII2023

Ethical concerns have been a core issue of Affective Computing even from the early years of the field. And the large diversity of research done in the Affective Computing community today also means that there is a large diversity of ethical issues. *Not every project will encounter every issue, but almost all projects will encounter some issues.*

Starting in 2022, the organizers of Affective Computing and Intelligent Interaction (ACII) have made writing an Ethical Impact Statement a mandatory part of the submissions process to be filled in on the EasyChair submission portal. **From 2023 onwards, authors of submissions to ACII will also be required to include an Ethical Impact Statement <u>as part of their paper submission</u>.**

This document will complement the ACII2023 submission instructions, and discuss how to write a thoughtful Ethical Impact Statement. This is based on the main themes identified from Ethical Impact Statement submissions to ACII2022, and will clarify some common misconceptions.

Quick links:
- ACII2023 Submission Instructions
- Previous years' instructions: links to ACII2022 submission instructions and ACII2022 Ethical Impact Statement FAQ.

Table of Contents:

Further Reading and References from other similar conferences. We acknowledge inspiration from some of these guidelines:
- NeurIPS 2021 Ethical Guidelines, NeurIPS Current Guidelines
- NLP: NAACL 2021 Ethics FAQ, EMNLP 2021 Ethics FAQ, ACL-IJCNLP 2021 Ethics FAQ
- A Guide to Writing the NeurIPS Impact Statement (on Medium)

Types of Ethical Issues

Affecting Computing research has always been associated with a host of ethical issues. With the good that our technology can achieve, we should also always be mindful of potential harm and risk of harm. Thus, it is important that authors of ACII submissions give careful thought as to the potential impact of their work and suggest means to mitigate the risks. In some cases, it may not be possible to draw a line between what is ethical and what is unethical, because ethical norms differ by society and discipline, but it is important nonetheless to have the discussion. **We encourage authors to view the Ethical Impact Statement as an opportunity to welcome discussion and to participate in a larger conversation about the direction of our field**, rather than as being "defensive" and judgmental**.**

ACII is a diverse community that welcomes a wide range of work: from theoretical (philosophy/psychology/technology studies) work, human subjects work, computational modeling and machine learning work, systems development work, to hardware development like wearables and other sensors. We believe there are important ethical issues for almost all the work done in ACII (except a very small minority of purely theoretical work), although **there are different ethical considerations for different types of projects.**

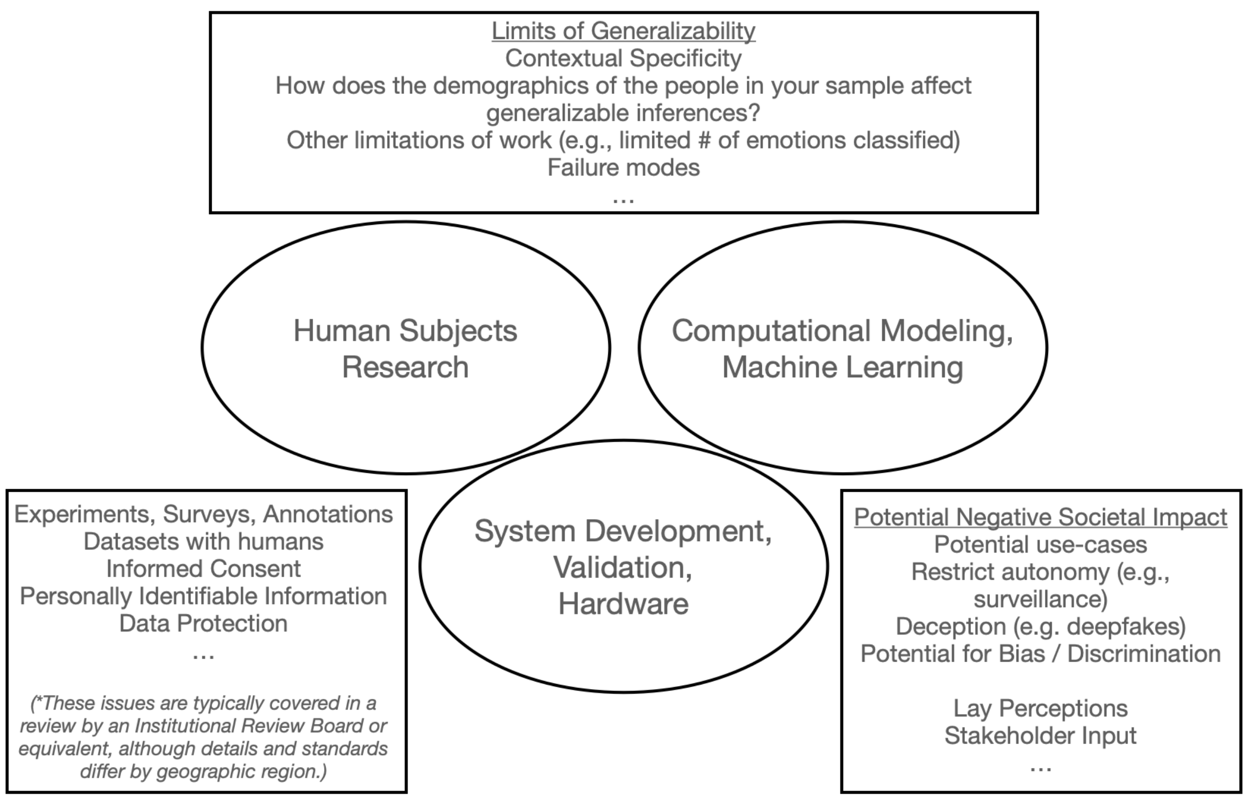

Fig. 1: Circles indicate broad classes of methodology in Affective Computing, including Human Subjects Research, Computational Modelling/Machine Learning, and System Development work, including hardware development. Research projects may span multiple methodologies. The three boxes correspond to three "themes", or collections of issues that could be relevant. Lists of issues are illustrative and non-exhaustive.

We have identified three broad themes of issues that concerns Affective Computing research (See Fig. 1), and authors are encouraged to read and reflect on the breadth of issues that may be relevant to their work.

<u>Issues related to human subjects</u>

This includes studies that run surveys and experiments with human participants. It also includes studies that propose the collection of new datasets that include people—even if the datasets are "public" information (e.g., from the Internet).

Here are some example issues for surveys and experiments:

- Was there ethics oversight in the collection of such data? For example: Was there Institutional Review Board (IRB) approval?

  Many institutions require such research to be approved by an Institutional Review Board or equivalent, BEFORE the research is carried out.
  *If the research was approved by an oversight board, details of such approval should be mentioned (*please take note that such disclosures in initial submissions have to be anonymized).*
  We recognize that standards will differ by geographic region, and some authors may not have a requirement in their region to seek ethics oversight.

- Did participants give informed consent? What were the steps taken to protect participant privacy?

If the paper includes a dataset,

- Does the dataset contain **personally identifiable information**?
  Or can people's identity or other information be deduced from the data (e.g., raw speech recordings)?
- Did the participants in the dataset give consent for their information to be used in the dataset?
- Does the dataset contain potentially offensive content? (This is also relevant if other people, such as human annotators, were exposed to the dataset.)

<u>Issues related to potential negative societal impact</u>

It is important to think about downstream use-cases even if the work done in the paper is "upstream" research, or a "proof of concept". Researchers that develop such technology have a moral obligation to think about potential misuses.

- <u>Potential negative applications:</u> Can the research or technology described be used in applications that limit human rights or impact people's livelihoods? For example, surveillance: can a state or company use the technology to monitor the emotions of its citizens and employees, perhaps against their consent and in ways that take away from their freedom and autonomy?

- Deceptive applications: Can the research or technology be used to deceive people? (e.g., deepfakes). What steps could be taken to prevent this?

- Potential for bias / discrimination: Does the research or technology contain bias against certain groups of people that could result in discrimination? Will it exacerbate already-existing biases (e.g., will it perpetuate gender or racial bias?)?

- Risks to privacy: What are the privacy considerations that should be taken into account for the discussed research, or downstream applications of the research?

- Lay perceptions: What are some legitimate ethical concerns that the general public could have about the research?

- Stakeholder input: Did you consult all the stakeholders involved? For example, if you are building a system meant to benefit a particular group (e.g., with a specific disorder), did you consult with people in that group to find out their concerns (e.g., about privacy, mis-use, whether they even need, want, or will use the system)? Without such input, systems may also be developed in a way that may benefit a group in one way but ends up harming them in other ways (e.g., by restricting autonomy, making them feel "de-humanized", etc).

Issues related to limits of generalizability

What are the important "caveats" that need to be highlighted about the research findings, with regards to how it will generalize to other contexts?

- Generalizability: How generalizable will the results of the research be, especially when considering extensive psychological research on the cultural variability of emotions, facial expressions, and other aspects of affect?

  How does the demographic makeup of the model's training dataset (e.g., trained on college students, or a specific racial group) limit generalizations to other populations?

  Are there enough details of the dataset used (e.g., demographic characteristics of people in the dataset) that can be used by readers to assess generalizability? Dataset papers should also consider suggestions like Datasheets (Gebru et al., 2021) to provide comprehensive documentation, and papers that propose models should consider Model Cards [paper]. These are proposals to improve the documentation of our datasets and models to be more transparent about what is in them and what is not in them (much like nutrition labels!)

  Are there potential biases in the dataset collection or annotation that might limit generalizability?

Note that in Psychology, there has been a recent movement to add "Constraints on Generalizability [non-paywall PDF]" sections in journal papers.

- Contextual Specificity: How sensitive is the research to contextual factors? For example, unimodal emotion recognition, by definition, ignores context: are there important blind spots in such work?

  Does the research include considerations about specific contexts (e.g., affective tutors in a classroom; personalization due to individual differences or cultural differences)? If the work is "generic", e.g., a generic facial expression classifier, what are the considerations about generalizing to particular contexts (there has been a lot written about the importance of context in facial expression recognition).

- Other limitations: If you are building an emotion classifier, what emotions are considered by the model? What types of representations are used?

  There are also often limitations of the theory, which may make simplifying assumptions that may not capture some aspects of real emotions. There are also often limitations of the measurements themselves. Self-reports may be biased for many reasons. Other-reports (such as using crowdsourced workers to annotate a dataset) could also be biased based on the demographics of workers, what is the contextual information they are told about the dataset, etc. Sensor data limitations and inferences from such data should be discussed.

- Biases: Could the technology have hidden biases? For example, is the dataset that it is trained on representative enough for the intended use-case? Could there be systematic issues or blindspots in the dataset / annotation / model training that could influence the predictions made by the technology? What could the consequences be?

- Failure modes: What are the limits of the technology; when do you expect it to fail, and what would happen in the case of failure? (E.g., if the technology is used to predict pain for a healthcare application, what may happen when it fails?)

Other Possible Issues

- For studies that involve lots of compute time and power, it may be prudent to also think about the potential energy (and carbon) costs to assess the impact of the work on the environment. Minimally, studies should report compute time / hardware (e.g., "Experiments were run on a NVIDIA Tesla T4 GPU and took 168 hours to complete"), which are also helpful for scientific reproducibility.

For ACII2023, we expect submissions to include discussion of the above issues, where applicable, including potential harms, mis-use, and other concerns. If the risks are inherent to some design choice that the authors made, authors should elaborate on the rationale for their decisions. When applicable, we also expect authors to include a discussion about any steps the authors have taken to mitigate such risks, or recommendations for future researchers who build upon such work.

A note on data handling: We encourage authors to adhere to best *ethical* practices (e.g., de-identifying raw data as soon as practically feasible, doing analysis only on de-identified data). Orthogonal to ethical considerations, we also encourage authors to adhere to best *data transparency* practices, including making one's data available for scientific reproducibility (e.g., "open science" and data sharing). We realize that these sometimes conflict, but whenever possible, we encourage that individual research groups take steps to release as much of their data as possible to allow reproducibility (including future re-analysis, meta-analysis, etc) while respecting data protection laws and protecting participants' privacy. If you plan to release your data, we also strongly encourage making this clear to your participants in the informed consent process (this is current best practice in behavioral fields).

		Possible issues related to:	
Hypothetical examples	Human Subjects	Potential Negative Social Impact	Limits of Genealizability
An experiment that induces emotional states and measures participant behavior via surveys and sensors.	IRB; informed consent; protecting privacy		Note demographic and other characteristics of human subjects sample, and if those characteristics limit inferences? Limits of sensors?
Development of a machine learning model that predicts human emotions from facial expressions, using publicly-available datasets		What are some negative applications in which this model can be used? How can we mitigate this risk? What are some biases that could be present in the model?	What are the contexts in which the training datasets are collected? What is the intended context that the authors have for deployment, and do they match?
Development and validation of a system designed to interact with depressed patients to improve their mental health	IRB, informed consent; protecting privacy	Did you consult with mental health professionals and/or potential target users to solicit their input? What are laypeople's perceptions of the risks/harms of the system?	What were the specifics of the context in which the system was tested? How sensitive is the system to contextual factors?

Table 1: Research at ACII is diverse, and not every paper will encounter every issue. There may also be issues that fall outside of these broad themes.

Top 10 Misconceptions, Explained

The ACII community is growing and learning together about approaching and discussing ethical issues in our research. In this section, we want to address the "top 10 misconceptions" that came up in discussions with community members and in the ACII2022 submissions.

1. ***"There are no ethical concerns"*** */ "This study does not have potential negative applications".*

Affective Computing concerns people, and in particular their emotions, which makes a lot of our research more "sensitive". Thus, it is extremely likely that there will be some concerns, if not many concerns. We encourage authors to read about the broad scope of ethical issues listed above, and also have conversations with colleagues and other members of the ACII community.

2. *"I am not doing social science work.* ***<u>I am doing 'technical' work</u>****. Therefore there are no issues".*

It is true that the most commonly-known ethical issues in research concern human subjects. But as we discussed above, there are many other types of ethical issues, some of which concern even the technical work that we do in ACII.

What has changed in the recent few years, especially in AI-related fields that includes affective computing, is a growing awareness that even "technical" work have ethical implications, ranging from negative societal impacts to issues with generalizability, bias, and more. That is why more AI communities are having a greater discussion of AI Ethics, and why the ACII community is committed to encouraging greater awareness.

3. *"I used a* ***<u>publicly available dataset</u>*** *to build my model, therefore there are no ethical issues".*

Using a previously-published dataset does not *de facto* satisfy all ethical concerns. Authors needs to reflect on the ethical aspects of their *new work* using that dataset.

For example, consider a facial expression recognition paper that uses previously-published FER datasets to develop better models which could potentially be used for surveillance; the authors should then discuss potential negative impacts as well as limits of generalizability.

4. *"We are working on a* ***Proof of Concept*** *and this work is not intended for applications as is".* Relatedly, *"We aim to explore such implications in our future work".*

Many researchers working on "upstream" work, "proof of concept" work, or "low Technology Readiness Level (TRL)" work, may feel that it is premature to think about downstream applications of their research. However, recent years have shown that the pace of technological development, especially in AI, is accelerating, and the time from POC to applications is getting shorter. Thus, it is imperative that researchers reflect on downstream applications and broader impacts of their work, now even in the "early" stages.

Relatedly, we also think that "aiming to explore such implications in future work" is an excellent intention, but that authors should still take some time now to reflect on possible ethical issues. The pace of technological development is so fast that we cannot wait for "future work" to think about the impact of today's research. **We cannot procrastinate ethics.**

5. *"We do not think our work will be used in critical applications"*

Unfortunately, many of us cannot predict the future, and it is impossible to guarantee that one's work will not be used in critical applications. Authors should be very cognizant and transparent, especially about the limits of generalzability of their work, to ensure that future work that builds upon this research—if ever used in real applications that affect people—are aware of such limitations and account for them.

6. *"We are doing emotion recognition **with good intentions.**"*

That's great, and we respect that. However, just because the authors of the current work have good intentions does not mean that downstream applications inspired by the work (which could be done by other researchers, or even by the same authors in the future) cannot be mis-used. Indeed, it is not commonly the case that researchers "set out" with the intention of developing technology for harmful applications, yet such harmful applications still exist.

Having good intentions in the research work does not absolve one from the need to think about possible harms and societal impacts. Authors should reflect on if and how they could mitigate future risks.

7. Relying on external guidelines and reviews completely prevents ethical concerns, e.g., ***"I have IRB approval**, there are no ethical issues."*; *"My study adheres to GDPR"*.

Research studies that receive Institutional Review Board (IRB) review and approval is a good start for addressing ethical concerns, but IRB review is often scoped on protecting participant rights. IRB review ensures that studies with human subjects are ethically conducted (informed consent, etc). Such reviews do not provide guidance or frameworks for how the results of the study will be used.

Similarly, regulations such as the European Union's General Data Protection Regulation (GDPR) ensure that the data collected is secure and used responsibly. This is more nuanced, but *meeting legal requirements does not necessarily mean that all ethical concerns are satisfied*.

For example, these external guidelines and reviews do not consider the implications of the work beyond the research study (i.e., potential negative social impact) or the limits of generalizability.

Thus, these guidelines are a great guide, but authors are encouraged to reflect beyond the "minimum" imposed by regulations and beyond the scope of such regulations.

8. Removing **personally identifiable information** prevents all ethical concerns

Removing personally identifiable information is a great start, but it is not enough to reduce the potential sources of user harms such as reidentification. Reidentification is the process of linking anonymous data to a specific individual, and can occur when an adversary uses a combination of public and private data sources to gain access to an individual's identity. Thus, even after removing personal identifiers such as name, address, and social security number, it is still possible for someone to link the data back to an individual. To reduce the potential for reidentification, data could be further anonymized, such as removing any unique patterns or attributes from the data that could be used to identify an individual. In addition, the data could be encrypted and stored securely, and access must be restricted to authorized personnel.

9. Pasting a EULA (end-user license agreement) or privacy policy. Or offering written committments like "we will use the data only for scientific and non-commercial purposes"

If your team has a EULA/privacy policy, or are willing to make committments like not using the data for scientific and non-commercial purposes, that is commendable. But this misses the point of the Ethical Impact Statement, which is to reflect on, for example, potential mis-uses of the work. Saying you won't mis-use the work does not mean the work could not be mis-used in the future (This is related to Misconception #6).

10. Self-reports or other-annotations are accurate assessments of emotions

Human annotations such as self-reports or external coders (other-reports) have long been the gold standard in emotion research. For example, many emotion recognition datasets are labelled by external raters who are asked to rate how the individuals in the stimuli are feeling. However, such ratings may contain bias (e.g., due to different cultural conceptualizations of emotion, or other forms of cultural bias) and other factors that may affect the validity of such ratings—and at the end of the day, they are just somebody's best guess. In addition, the accuracy of annotations is heavily dependent on the context of the application. Many scholars have also pointed out the difference between recognizing a facial expression/facial movements, versus inferring the underlying emotions that someone is feeling (e.g., Barrett et al., 2019). Emotion recognition models should avoid treating annotations as the ground truth of internally felt emotions, especially when ignoring the context. Rather, just like any other measurement (e.g., with sensors), we should be discussing the limitations of the measurement and inferences that we can draw from such data. (As an analogy, we do not take psychophysiological data to be "readouts" of an underlying emotion, and we are careful about the limited inferences that we can make from such data.) This is important to help prevent potential risks associated with downstream applications of such models.

Frequently Asked Questions

Q: My research is very "upstream" / on "basic science" and far from applications. Do I really need to consider the ethical impacts of my research?

A: Yes! Especially for AI research, the turnaround from "basic research" to applications can be as short as a few months. Everyone doing AI research should be concerned about the downstream impact of their work, no matter how "far" the researcher thinks the research is from applications. See "Misconception #4" and "#5".

Q: I am working on technology that I want to use for positive applications. I guess it could be used by others for negative applications as well, but that is out of my control!

A: This is exactly the point: AI technology is a tool that could be used in many different ways, some more questionable than others. The researchers that develop such technology have a moral obligation to think about potential misuses. See "Misconception #6".

Q: But my paper has no ethical issues (e.g., because it is a theoretical paper / …) ! How should I put this into my paper?

A: Please read Misconception "#1" first.
If you feel that the Ethical Impact Statement does not apply to your paper, you may write something to that extent in the paper (e.g., "This is a purely theoretical paper that considers XYZ, and we believe there are no ethical issues to discuss at the moment.").
But we encourage authors to nevertheless spend some time reflecting. Even theoretical work may have important downstream implications.

Q: Can a paper be rejected based on the Ethical Impact Statement?

A: We will be providing instructions to reviewers to also review the Ethical Impact Statement to the best of their ability (e.g., judging whether it is sound). Reviewers will have the option to flag concerns up to the SPC members and Program Chairs, who will consult with the Ethics Committee. These will be discussed on a case-by-case basis. If necessary, we may ask for revisions or additional discussion to be added to camera-ready papers. The ACII Organizing Committee reserves the right to reject papers that have egregious violations.

Q: How long should the Ethical Impact Statement be?

A: It is difficult to offer a length prescription, as different papers will have different issues. *(See here for a Medium post for the NeurIPS Impact Statement, which includes several examples from real papers.).*

List of Contributors, Acknowledgements, References

Desmond Ong (Ethics Chair, ACII2023 & Program Chair, ACII2022)
Javier Hernandez
Rosalind Picard
Fabien Ringeval
Jeffrey Girard
Agata Lapedriza
Akane Sano

We would like to thank the following people who provided feedback on earlier versions of this document:
Rachael Jack
Shiro Kumano

References

Cowie, R. (2015). Ethical issues in affective computing. *The Oxford Handbook of Affective Computing*, 334-348.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86-92.

Hernandez, J., Lovejoy, J., McDuff, D., Suh, J., O'Brien, T., Sethumadhavan, A., Greene, G., Picard, R., & Czerwinski, M. (2021). Guidelines for assessing and minimizing risks of emotion recognition applications. In *2021 IEEE 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*.

Ong, D. C. (2021). An ethical framework for guiding the development of affectively-aware artificial intelligence. In *2021 IEEE 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*.

Stark, L., & Hoey, J. (2021). The ethics of emotion in artificial intelligence systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 782-793.

[*There is an ongoing special issue on Ethics in IEEE Transactions on Affective Computing; if there are relevant articles published there we will add them here.]